

Institute for Research on Poverty
Discussion Paper no. 1224-01

**Measuring Employment and Income for Low-Income Populations
with Administrative and Survey Data**

V. Joseph Hotz
Department of Economics
UCLA
E-mail: hotz@ucla.edu

John Karl Scholz
Institute for Research on Poverty
and
Department of Economics
University of Wisconsin–Madison
E-mail: jkscholz@facstaff.wisc.edu

February 2001

Printed, with permission, from the Workshop on Data Collection on Low-Income and Welfare Populations, December 1999. Panel on Data and Methods for Measuring the Effect of Changes in Social Welfare Programs, Committee on National Statistics, National Research Council. Copyright 2000 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, DC. We thank Howard Bloom, Paul Burgess, Martin David, George Falco, Julie Isaacs, Rebecca Maynard, Robert Moffitt, Anne Polivka, Jeff Smith, Shelly Ver Ploeg, Dan Weinberg, and conference participants for helpful comments on this project.

IRP publications (discussion papers, special reports, and the newsletter *Focus*) are available on the Internet. The IRP Web site can be accessed at the following address: <http://www.ssc.wisc.edu/irp/>

Abstract

We discuss the strengths and weaknesses of income and employment data in national surveys, in unemployment insurance (UI) wage records, and in tax returns. The CPS, SIPP, NLS, and PSID surveys provide valuable information on the behavior of the low-income population. They have broad and fairly accurate measures of income for national samples, and their focus on families as the unit of analysis and their ease of access greatly enhance their value.

The value of these data sets for evaluating welfare reform is severely limited, however. With the devolution of responsibility for TANF, the CPS and SIPP sampling frames and sample sizes mean that, at best, they can be only supplementary data sources for understanding the effects of welfare reform at the state and local levels. The apparent decline in program coverage in the CPS is also worrisome.

UI data are available at the state level and can be matched to individuals in existing samples at relatively low cost. It is straightforward to do follow-up analyses on income and employment for workers who remain in the state, and UI data are timely. However, earnings are available only for individuals, while changes in family composition upon exit from welfare have been shown to have a large bearing on economic well-being. UI data do not allow us to track these changes. There also appears to be a substantial problem with some workers being classified as independent contractors and hence not entering the UI system. Overall gaps in coverage appear to be at least 13 percent and may be significantly higher. Even when wages are reported, there is some evidence that they are understated by a significant amount.

We also present evidence on the degree to which tax data can be used to understand the incomes and employment of low-skilled workers. The paper concludes with brief recommendations for future research that might help fill some of the gaps we have identified.

Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data

1. INTRODUCTION

With passage of the Personal Responsibility and Work Opportunities Reconciliation Act (PRWORA) of 1996 and the expansions of the Earned Income Tax Credit (EITC) over the last decade, increasing attention has been paid to the employment experiences, labor market earnings, and transfer income received by disadvantaged individuals and households in the United States. This attention, prompted by explicit performance goals in PRWORA and implicit goals of the EITC expansions, focuses on whether low-income households can achieve self-sufficiency without resorting to TANF or other public assistance programs. Income and employment levels are only partial indicators of the well-being of households, but they continue to be ones most often used to assess the consequences, intended and unintended, of welfare reform.

More broadly, good measures of income and employment for low-income families are necessary to (1) assess the well-being and labor market attachment of low-income and welfare populations at the national, state, and local levels, (2) evaluate welfare reform and learn the effects of specific policies, such as time limits and sanctions, and (3) meet reporting requirements under TANF and aid in the administration of welfare programs.

There are two sources of data—survey and administrative—for measuring employment and incomes of the disadvantaged. Surveys have been the mainstay of evaluating welfare programs and of monitoring changes in income and employment for decades. These include national surveys—such as the U.S. Census of Population, the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the National Longitudinal Surveys (NLS), and the Panel Study of Income Dynamics (PSID)—and more specialized surveys that gather data for targeted groups (such as current or former

welfare recipients) and at the state or local level.¹ Although survey data continue to be important, administrative data sources to measure income and employment have become increasingly valuable over the last 30 years. Data on wages and salaries from state unemployment insurance (UI) systems, for example, have been used to measure the earnings and employment of individuals participating in state AFDC (TANF) programs, manpower training, and other social programs. Data on earnings (and employment) from Social Security Administration (SSA) records have been linked with the records of welfare and social program participants.

What type of data one uses to measure income and employment among current and past welfare participants and welfare-eligible households may have important consequences for implementing and evaluating recent welfare reforms. Recent debates between the states and the federal government, for example, over employment targets and associated sanctions mandated under PRWORA hinged crucially on exactly how the fraction of a state's caseload that is employed would be measured. Furthermore, the conclusions of several recent assessments of the impacts of welfare reform and caseload decline appear to depend on how income and employment of welfare leavers and welfare-eligible populations are measured.²

In this paper we assess the strengths and weaknesses of using survey or administrative data to measure the employment and income of low-income populations. We review a number of studies, most of which have been conducted in the last 10–15 years,³ that assess the comparability of income and employment measures derived from surveys and administrative records. Clearly, the primary criterion for evaluating data sources is their *accuracy* or *reliability*. Ideally, one would compare the income and

¹Often these samples are gathered in the context of evaluations of specific welfare or training programs.

²See, for example, studies by Primus et al. (1999), Cancian et al. (1999), and Rolston (1999) for a flavor of how this debate hinges on measurement issues.

³Several earlier studies compared employment measures for low-income populations across alternative data sources, most notably the study by Greenberg and Halsey (1983) with data from the SIME/DIME experiments. Given changes over time in such things as UI coverage and response rates in surveys, we focus on the most recent studies available to maximize the relevance of our findings for the measurement of these outcomes for current and future studies.

employment measures derived from either surveys or administrative data sources with their true values to determine which source of data is the most accurate.

Unfortunately, this ideal is rarely achieved. One seldom, if ever, has access to the true values for any outcome at the individual level. At best, one can determine only the *relative differences* in measures of a particular outcome across data sources. In this paper, we try to summarize the evidence on these relative differences and the state of knowledge as to why they differ. These studies point to several important dimensions along which surveys and administrative records differ and, as such, are likely to account for some, if not all, of the differences in the measures of income and employment derived from each. These include:

- *Population Coverage:* Surveys generally sample the population while administrative data typically cover the population of individuals or households that are enrolled in some program or are subject to some reporting requirement. In each case issues arise about the sizes of samples at state or substate levels and sample designs that may limit the issues that can be examined.
- *Reporting Units:* Different data sources focus on individuals, households, tax-filing units, or program case units. Differences in reporting units hinder the ability to move across data sources to obtain measures of income and complicate efforts to evaluate the differential quality of income data across data sets. Furthermore, differences in reporting units may have important consequences for the comprehensiveness of income measures, an issue especially relevant when attempting to assess the well-being, and changes in the well-being, of disadvantaged populations.
- *Sources of Income:* Data sources differ in the breadth of the sources of individual or household income they collect. Surveys such as the CPS, and especially the SIPP, attempt to gather a comprehensive set of income elements, including labor earnings, cash benefits derived from social programs, and income from assets. In contrast, administrative data sources often only contain information on a single type of income (as in the case of UI earnings) or only those sources of income needed for the purposes of a particular record-keeping system.
- *Measurement Error:* Different data sources may be subject to different sources of measurement problems, including item nonresponse, imputation error, and measurement error with respect to employment and income (by source). Furthermore, issues such as locating respondents, respondent refusals, and sample attrition are important in conducting surveys on low-income populations.
- *Incentives Associated with Data Gathering Mechanisms:* Data sources also may differ with respect to the incentives associated with the gathering of information. In the case of surveys, respondents' cooperation may depend on a comparison of the financial remuneration for a survey with the respondent "burden" associated with completing it. In the case of administrative data, the incentives relate to the administrative functions and purposes for which the information is

obtained. What is important is attempting to anticipate the potential for and likelihood of biases in measures of income and employment that may result from such incentives.

The importance of various strengths and weaknesses of different data sources for measuring employment and income generally will depend on the purpose to which these measures are put. We note five considerations. First, when conducting an experimental evaluation of a program, the most important criterion for judging data sources is whether they yield different quality estimates of program impact, which generally depends on *differences* in income (employment) between treatment and control groups. In this case, similar errors in measuring the level of income between treatment and control groups could have little effect on the overall evaluation results. Alternatively, if the objective is to describe what happened to households that left welfare, researchers will be interested in the average *levels* of postwelfare earnings (or employment). We discuss results of a study by Kornfeld and Bloom (1999) in which unemployment insurance (UI) data appear to understate the level of income and employment of treatments and controls in an evaluation of the Job Training Partnership Act, but *differences* between the two groups appear to give accurate qualitative measures of program impacts. Depending on the question of interest, the UI data may be suitable or badly biased.

Second, surveys, and possibly tax return data, can provide information on both individual and family resources while UI data provide information on individual outcomes. When assessing the well-being of case units who leave welfare, we are often interested in knowing the resources available to the family. When thinking about the effects of a specific training program, we often are interested in the effects on the individual who received training.

Third, data sets differ in their usefulness in measuring outcomes over time versus at a point in time. UI data, for example, make it relatively straightforward to examine employment and earnings over time, while it is impossible to do this with surveys unless they have a longitudinal design. Even with longitudinal designs, there are problems of varying data accuracy associated with length of recall.

Fourth, sample frames differ between administrative data and surveys. Researchers cannot use administrative data from AFDC/TANF programs, for example, to examine program take-up decisions since the data only cover families that already receive benefits. Surveys, on the other hand, generally have representative rather than targeted or “choice-based” samples.

Fifth, data sources are likely to have different *costs*. These include the costs of producing the data and implicit costs associated with gaining access. The issue of access is often an important consideration for certain sources of administrative data, particularly data from tax returns.

The remainder of this paper is organized as follows. We characterize the strengths and weaknesses of income and employment measures derived from surveys, with particular emphasis on national surveys, from UI wage records, and from tax returns. For each data source, we summarize the findings of studies that directly compare the income and employment measures derived from that source with measures derived from at least one other data source. We conclude by identifying the gaps in existing knowledge about the survey and administrative data sources for measuring income and employment for low-income and welfare-eligible populations. We offer some recommendations for future research that might help to close these gaps.

2. USING SURVEY DATA TO MEASURE EMPLOYMENT AND INCOME

In this section, we discuss the strengths and weaknesses of measuring income and employment status for low-income populations using survey data. Most of our analysis focuses on national surveys—the CPS and the SIPP in particular—because of the availability of several high-quality studies that compare their income and employment measures to other data sources. We also summarize available studies that assess income and employment measurement with more targeted surveys.

The key features of the national surveys for the purposes of this paper are summarized in Table 1.

TABLE 1
Key Features of Selected National Surveys That Report Employment and Income Status of Individuals and Households

Feature	Current Population Survey (CPS)	Survey of Income and Program Participation (SIPP)	Panel Study of Income Dynamics (PSID)	National Longitudinal Survey of Youth, 1979 (NLSY79)
Nationally representative sample?	Yes	Yes	Only at sample inception in 1968	No, but representative for cohorts covered at sample inception
Primary unit of analysis	Household	Household	Household	Individual
Longitudinal data?	No	Yes	Yes	Yes
Typical sample size	60,000 households	21,000 households	8,700 households	11,400 individuals
Capacity for state and local analysis	For all but small states	For large states only	Limited	Limited
Coverage of income sources	Broad	Very broad	Broad	Very broad
Accuracy of earnings data ¹	97%	92%	—	—
Accuracy of AFDC data ²			—	—
Timeliness of data	Several months	2+ years	2 year lag	1–2 year lag

¹For 1990; see Table 3.

²For 1990; see Tables 2 and 3.

2.1 Potential Strengths

The CPS and SIPP are vital data sets for understanding the functioning of low-wage labor markets and the effects of antipoverty programs. These data get high marks on many of the concerns mentioned in the introduction. They have a national sampling frame covering program participants and nonparticipants, thus making these data valuable for developing a broad perspective on developments in low-wage labor markets. An example of this type of study is Primus et al. (1999), which uses CPS data to show that AFDC/TANF and Food Stamp program participation rates declined considerably faster than poverty rates between 1993 and 1997. They further report that incomes of poor, single mothers fell between 1995 and 1997 (after rising between 1993 and 1995) and that the safety net is lifting fewer children from poverty than in the past. Some concerns with this study are mentioned below. Nonetheless, the CPS and SIPP are the only data sets that would allow analysts to address the important issues that Primus et al. examine on a national scale.

The other national data sets that have been used to analyze the employment and income status of low-income populations are the NLS (particularly the NLSY79) and the PSID. Both have the additional feature of being longitudinal surveys so that one can obtain information on earnings and employment status over time for the same person (and household).⁴ The PSID has surveyed, until very recently, its respondents and the “splitoffs” of initial respondent households on an annual basis since 1968. Similarly, the NLSY79 has, until 1994, conducted annual surveys of a random sample of individuals who were 14–21 years old in 1979. Both of these surveys gather detailed information on labor market earnings and employment status of respondents, earnings and some employment information of other adult household members, and some information on other sources of income, including income from various public assistance programs. One advantage of longitudinal data sets like the SIPP, PSID, and NLSY is that one

⁴Each wave of the SIPP is a longitudinal survey with between 2.5 and 4 years of data on the residents of a sample housing unit. Surveys to these respondents are conducted every 4 months.

can monitor the entry into and exit from welfare or other social programs and the factors related to welfare dynamics, including changes in earnings and family structure.

The CPS, SIPP, and PSID, in addition to having nationally representative samples, focus on households as the unit of analysis and include information on all adult household members.⁵ Given the general presumption that families pool resources, data sets that focus on families or households (and include information on cohabiting partners) are valuable. A calculation in Meyer and Cancian (1998) illustrates the usefulness of having data on family, as well as individual, incomes. Their study examines the economic well-being of women in the 5 years after leaving AFDC. They show that in the first year of exit from AFDC, 79.0 percent of the women have incomes below the poverty line, but when family income is considered, just 55.5 percent have income below the (correspondingly larger) poverty line. After 5 years, 64.2 percent of the women, but only 40.5 percent of the broader family unit, still had incomes below the poverty line.

The nationally representative surveys, especially the SIPP, provide information on multiple sources of income, either through separate questions or prompting of specific income sources. By asking specific questions about, for example, welfare receipt or Food Stamps, the data identify participants and (eligible) nonparticipants, so the data can be used to study program entry effects.

The national surveys also measure income and employment in a comparable fashion both over time and across geographical locations, though in January 1994 the way that earnings information was elicited in the CPS was changed (Polivka, 1997).⁶

⁵The NLSY79 focuses on the original respondent but it gathers a considerable amount of information on the respondent's spouse and/or cohabiting partner.

⁶Previously, earnings had to be reported in weekly amounts, and amounts over \$2,000 per week were truncated. Now earnings can be reported over any interval and the data (to the Bureau of Labor Statistics) are not truncated. Studies using repeated cross sections of the CPS spanning 1994 risk misinterpreting results if they fail to account for the redesign. Polivka provides adjustment factors for earnings (at the 10th percentile, median, and 90th percentile) reported before 1994 to make the series comparable. She also shows that "top-coded" values imputed using a Pareto distribution do a good job of fitting the distribution of data that are not top-coded.

Another strength of the nationally representative surveys is that questions can be modified to reflect changing circumstances. For example, the U.S. Census Bureau periodically conducts cognitive interviews of respondents to the CPS to assess how they respond to different CPS income- and welfare-related questions. Such studies help determine which of the CPS questions are confusing and how respondents interpret questions. Results from these cognitive interviews are used to improve the way questions are asked, with the goal of improving the quality of the data on such key variables as income and program participation.⁷ Typically, this sort of sophisticated assessment can only be done on large-scale, national surveys.

To summarize, there are several potential strengths of using survey data to measure income and employment.

- Surveys can provide representative samples for specific populations and generally include data for other family members.
- Surveys typically provide demographic data and data on other characteristics of households (such as educational attainment). They also may gather detailed information on many distinct income sources.
- National surveys provide consistent information across states and localities.
- Surveys can be flexible, so their developers can control what information is collected about income and employment, and this information can be improved over time.

2.2 Potential Weaknesses

Three general concerns arise with the nationally representative surveys that keep them from being the solution, or “core” data, for understanding the effects of welfare reform. The most important issue is that sample sizes and sampling frames prevent these data from being used to examine certain subpopulations of interest, such as welfare recipients in a particular state (perhaps with the exception of large states, such as California, New York, and Texas). A distinguishing feature of welfare reform is that program responsibility now rests largely with states and even counties within a state. The nationally

⁷See Bogen et al. (1997), Bogen (1998), and Loomis and Rothgeb (1999).

representative data sets do not have sample designs and sample sizes that allow analysts to examine behavior at a level corresponding to where program decisions are being made.

Second, systematic changes in the coverage of low-income populations seem to have occurred in the CPS. Studies have found that AFDC and Food Stamp benefits and the number of recipients have declined in the CPS over time relative to estimates of participants from administrative records. This issue of coverage is a serious concern for studies that employ the CPS for measuring the income of welfare populations.⁸ In Table 2, we reproduce comparisons of aggregate AFDC/TANF and Food Stamp benefits between CPS and administrative data sources from Primus et al. (1999). It shows a sharp decline between 1990 and 1997 in the percentage of AFDC/TANF and Food Stamp benefits reported in the CPS compared with amounts reported in administrative data.⁹ The reduction in coverage of AFDC/TANF (or family assistance) benefits also is consistent with Roemer's (2000) calculations from the CPS (see Table 3b) for 1990 through 1996. Interestingly, the apparent decline in AFDC/TANF coverage does not show up in the SIPP, though SIPP appears to capture only about three-quarters of aggregate benefits.

Polivka (1998) compares the monthly average number of AFDC recipients in the March CPS to the monthly average reported to the Department of Health and Human Services (prior to quality control). She finds there has been a modest decrease in the proportion of total months on AFDC as measured in the CPS. The percentage of the CPS estimated count to the administrative count (excluding Guam, the Virgin Islands, and Puerto Rico) is 83.0 (1989), 86.7 (1990), 86.0 (1991), 82.5 (1992), 84.2 (1993), 78.5 (1994), 75.5 (1995), and 79.6 (1996). The timing of the drop in the percentage corresponds to changes in the March CPS survey instrument. Taken together, the Primus et al. (1999) and Polivka (1998) results

⁸Roemer (1999) suggests that the reduction in coverage could be related to PRWORA. The March 1997 CPS did not use state-specific labels for TANF benefits in 14 states that had abolished AFDC. Benefit estimates were 4.5 percentage points lower than the benchmark in states that had abolished AFDC than in states that had not. The delivery mechanism of benefits in some circumstances (for example, through employers), an enhanced sense of stigma, and caseload reductions that exacerbate recall errors may also contribute to underreporting.

⁹Primus et al. adjust the CPS data proportionately to account for the decline in benefits over time, but the value of this adjustment depends on the patterns of discrepancies in the data. Unfortunately, we know little about the factors associated with the underrepresentation of program participants in the CPS or the SIPP.

TABLE 2
AFDC/TANF and Food Stamp Aggregate Benefits Paid Based on Administrative Data
Compared with Estimates from Current Population Survey (calendar year, in billions of dollars)

	AFDC/TANF Benefits			Food Stamp Benefits		
	CPS Data	Administrative Data	Ratio (%)	CPS Data	Administrative Data	Ratio (%)
1990	14.259	18.855	75.6	10.335	13.556	76.2
1991	15.554	20.804	74.8	12.373	16.551	74.8
1992	15.362	22.258	69.0	13.394	20.014	66.9
1993	17.540	22.307	78.6	15.010	22.253	67.5
1994	17.145	22.753	75.4	15.317	22.701	67.5
1995	15.725	21.524	73.1	14.542	22.712	64.0
1996	13.494	19.710	68.5	14.195	22.440	63.3
1997	10.004	15.893	62.9	12.274	19.570	62.7

Source: Primus et al. (1999), p. 65, which in turn gives the source as HHS and USDA administrative records, CBPP tabulations of CPS data.

suggest that the decline in benefits reported in the CPS comes about from both a reduction in the coverage of families receiving AFDC and an underrepresentation of benefits conditional on receipt, though the second factor seems quantitatively more important than the first.

The third potential weakness of national surveys is that there is little or no “cost” to respondents of misreporting of income, employment, or other circumstances.¹⁰

Some specific potential weaknesses associated with the PSID and NLSY79 are of potential relevance for obtaining information on the income and employment status of low-income populations. Most notable is the fact that they are not, by design, representative of the general population over time. Both data sets began with samples that were representative of their targeted groups—young adults in the case of the NLSY79 and the national population as of 1968 in the case of the PSID—but are not designed to be representative of the national population, or even of the age group covered in the NLY79, in subsequent years. As such, this feature can result in biased measures of summary statistics on income and employment vis-à-vis the nation as a whole in more recent years.

The other feature of the NLSY79 and PSID relevant for assessing the income and employment status of low-income population is their respective sample sizes. The original sample for the NLSY79 was 12,686 young men and women; approximately 90 percent of the original sample remains today. The original sample in the PSID was 5,000 U.S. households in 1968 and, because of its growth due to accumulation of additional households through splitoffs from original households, it contained over 8,700 in 1995. These are not small sample sizes, but the sizes of low-income samples at a point in time are relatively small compared with both the CPS (which contains some 60,000 households at a point in time) and most waves of the SIPP (which, in its larger waves, contains data on 21,000 households). The sizes of

¹⁰Shroder and Martin (1996), for example, show that subsidized housing (broadly defined) is badly reported on surveys, including the American Housing Survey (and presumably SIPP). An underlying problem is that the phrase “public housing” means different things to different people, ranging from only projects to any kind of subsidized housing.

the low-income or welfare subsamples in the NLSY79 and PSID for even the largest states are generally too small to derive reliable measures on income and employment, let alone other outcomes.

To summarize, there are two primary, potential weaknesses with using national survey data to measure income and employment of low-income populations.

- Sample sizes in national surveys often are small for studies focusing on welfare or low-income populations, or that wish to examine specific targeted groups, such as current or former welfare recipients.
- The coverage of both recipients and benefits appears to be falling in national surveys.

2.3 Direct Assessments of Income and Employment Measures from Survey Data

Moore, Stinson, and Welniak (1997) conducted a general survey of income reporting in the CPS and SIPP, and Roemer (2000) assessed trends in SIPP and CPS income reporting between 1990 and 1996.¹¹ A central finding in both studies was that there is underreporting of many types of income in surveys. The reasons for this (and, hence, solutions in the design of effective surveys) are complex, however. The estimated magnitudes of CPS and SIPP underreporting for selected years are given in Tables 3a and 3b. (Note that differences may be the result of both flawed benchmarks and flawed surveys.)

Surveys of Income Reporting in the SIPP and CPS

The understatement of certain types of income, such as interest and dividend receipts, is probably not critical for low-income populations since low-income families typically receive small amounts of income from these sources. Based on the evidence presented in Tables 3a and 3b, it appears that wages and salaries are fairly accurately reported in the CPS, although less accurately in the SIPP. But Moore, Stinson, and Welniak (1997) note that 26.2 percent (35,205,000 out of 134,135,000 total weighted cases)

¹¹No comprehensive assessments of the quality of income and employment measurements for either the NLSY79 or the PSID have been undertaken. Roemer (1999) and Nelson et al. (1998) update the CPS calculations to 1996. Roemer (2000) also provides a nice discussion of adjustments that need to be made to compare aggregate SIPP and CPS totals to National Income and Product Account data.

TABLE 3a
Ratio of SIPP and CPS March Income Supplement Aggregate Income
Estimates to Independent Aggregate Income Estimates for 1984 and 1990

Source of Income	1984			1990		
	Independent Estimate (billions \$)	SIPP (%)	CPS (%)	Independent Estimate (billions \$)	SIPP (%)	CPS (%)
Employment Income:						
Wages and salaries	\$1,820.1	91.4	97.3	\$2,695.6	91.8	97.0
Self-employment	192.6	103.1	70.2	341.4	78.4	66.8
Asset Income:						
Interest	\$244.8	48.3	56.7	\$282.8	53.3	61.1
Dividends	59.3	65.9	51.8	126.3	46.1	31.3
Rents and royalties	19.4	211.3	95.4	44.1	102.9	87.8
Govt. Transfer Income:						
Social Security	\$160.5	96.2	91.9	\$225.5	98.3	93.0
Railroad retirement	5.6	96.4	71.4	6.9	95.7	66.7
SSI	9.9	88.9	84.8	13.6	94.9	89.0
AFDC	13.9	83.5	78.4	19.7	70.1	71.6
Other cash welfare	2.0	135.0	120.0	2.9	86.2	80.2
Unemployment insurance	16.3	76.1	74.8	17.7	84.2	80.2
Workers' compensation	14.1	56.7	48.2	14.6	86.3	94.5
Vets' pensions and comp.	13.9	82.0	59.7	13.8	84.1	77.5
Retirement Income:						
Private pensions	\$65.2	63.8	57.2	\$70.2	107.1	110.8
Federal employee pensions	20.3	98.0	84.7	30.4	73.4	82.6
Military retirement	15.6	105.1	98.1	20.4	92.2	89.2
S&L employee pensions	21.9	88.1	71.7	36.1	75.1	80.1
Miscellaneous Income:						
Alimony	\$2.7	100.0	81.5	\$2.5	116.0	124.0

Source: Adapted from Coder and Scoon-Rogers (1996).

TABLE 3b
Ratio of SIPP and CPS March Income Supplement Aggregate Income
Estimates to Independent Aggregate Income Estimates for 1990 and 1996

Source of Income	1990			1996		
	Independent Estimate (billions \$)	SIPP (%)	CPS (%)	Independent Estimate (billions \$)	SIPP (%)	CPS (%)
Employment Income:						
Wages and salaries	\$2,727.7	90.1	95.9	\$3,592.3	91.0	101.9
Self-employment	333.5	85.1	68.5	475.9	69.1	52.6
Asset Income:						
Interest	\$258.5	56.7	67.1	\$187.1	50.2	83.8
Dividends	96.8	65.8	40.9	129.4	51.0	59.4
Rents and royalties	45.6	113.1	85.0	76.2	82.0	58.6
Govt. Transfer Income:						
Social Security and railroad retirement	\$283.4	97.1	90.6	\$332.2	87.9	91.7
SSI	15.3	83.1	78.9	26.5	101.4	84.2
Family assistance	18.9	75.6	74.4	19.8	76.3	67.7
Other cash welfare	2.9	81.9	85.6	3.4	114.0	80.5
Unemployment insurance	17.9	77.5	79.9	21.6	69.4	81.6
Workers' compensation	15.4	67.8	89.5	17.0	71.7	62.7
Vets' pensions and comp.	14.5	83.1	73.9	17.8	72.9	89.6
Retirement Income:						
Private pensions	\$68.5	91.8	98.3	\$98.7	98.1	93.1
Federal employee pensions	30.5	75.9	82.7	38.8	75.6	80.8
Military retirement	21.4	87.4	85.6	28.3	101.6	58.2
S&L employee pensions	36.9	76.8	78.7	66.0	67.8	57.3

Source: From Roemer (2000). The Independent Estimates are the mean values of the implied independent estimates from SIPP and the CPS (from Tables 2a, 2b, 3a, and 3b in Roemer, 2000).

of the wage and salary “responses” in CPS surveys are imputed from cases where the respondent did not give an answer, replied “don’t know,” or refused to answer the question. They also report that 7 to 8 percent of households refuse to participate in the CPS, so imputations and imputation quality are clearly a critical element in survey quality.

The apparent accuracy of wage and salary reporting in Tables 3a and 3b does not fully resolve concerns that we have about data accuracy for low-income populations, since we do not know much about the characteristics of families that underreport their incomes. If, for example, most of the underreporting of income occurs among the disadvantaged, the findings of Moore, Stinson, and Welniak (1997) and Roemer (2000) on wage and salary reporting in the CPS and SIPP may be of little comfort. Roemer, for example, shows there are significantly more aggregate dollars reported below family income of \$25,000 in the SIPP than in the March CPS. He suggests that the SIPP does a better job than the CPS in capturing the incomes of low earners and a worse job capturing the incomes of high earners. Learning more about the nature of underreporting would appear to be a high priority for future research.

Matching Studies of Wage and Salary Income

Roemer (2000) examines the accuracy of CPS wage and salary reports by matching CPS data to IRS tax returns in selected years for the first half of the 1990s. The sample is limited to nonjoint returns and selected joint returns where each filer matches a March CPS person. The sample is further restricted to observations with no imputed wages in the CPS. He finds that in the middle of the income distribution (from \$15,000 to \$150,000) at least half the CPS and tax reports are within 10 percent of each other. Anywhere from 60 to 80 percent of the observations are within 15 percent of one another. Discrepancies appear much larger in the bottom and very top of the income distribution. Below \$10,000 and above \$150,000, at least half the observations have discrepancies exceeding 20 percent, and most are larger than that. Discrepancies are both positive and negative. As expected, however, CPS incomes tend to be larger than incomes reported on tax returns in the bottom of the income distribution, and CPS incomes tend to be smaller than incomes reported on tax returns in the top of the income distribution.

Beyond the cited studies, little recent work appears to have been done on the accuracy of the wage and salary income in the SIPP, CPS, or related national surveys.¹² The dates for the citations for American work on this topic (there also is one Canadian study) are 1958, 1970, and 1980, and in each case there seemed to be a small (on the order of 5 percent) incidence of nonreporting of wage and salary income.¹³ Coder (1992) compares a restricted set of SIPP households with tax data (married couples with valid Social Security numbers who file joint returns and have positive wage and salary income in either the SIPP or on tax returns) and finds a roughly 5 percent discrepancy in the existence of wage and salary income. Moore, Marquis, and Bogen (1996) examine a sample of SIPP households working for specific employers and find that respondents sometimes drop months of wage and salary receipt over a 4-month interview cycle, though virtually all accurately reported the presence of a job during the wave.

Several other studies assess the quality of income and earnings measurement based on matching survey data with various types of administrative data. Bound and Krueger (1991) match CPS data from 1978 and 1977 with SSA earnings records and find essentially zero net bias in CPS income reports for those whose incomes did not exceed the SSA's earnings maximum cutoff. In fact, more than 10 percent of the CPS sample matched their Social Security-reported earnings to the dollar, and 40 percent were within 2.5 percent. Rodgers, Brown, and Duncan (1993) examine wage records in the PSID for unionized men working full-time at an hourly rate in one specific durable goods manufacturing firm in 1983 and 1987. These authors examine three common measures of earnings: earnings from the previous week, from the previous year, and "usual" earnings. They find annual earnings to be fairly reliably reported, but this

¹²Abraham, Spletzer, and Stuart (1998) conclude, "There is some evidence that CPS hours worked are overreported, that this overreporting may have worsened over time.... Given the paucity of data on hours worked, we view our conclusions on this subject as suggestive rather than definitive" (p. 319).

¹³Moore, Stinson, and Welniak (1997) also provide a brief discussion of income data collected as part of the Gary Negative Income Tax Experiment (from the late 1960s and early 1970s). They note that the income data in this experiment, gathered through surveys of respondents, were not very reliable. In the Seattle and Denver Income Maintenance Experiments, there was evidence of statistically significant underreporting of wage and salary amounts, but the magnitude of underreporting was only between 2 and 4 percent, leading Halsey (1978) to conclude that they were not large enough to be economically important. The correlation between administrative records and reported values was 0.9, also indicating high reliability.

is less true for the other two measures. They also find that for each measure there is a tendency for workers with lower-than-average earnings to overreport and for workers with higher-than-average earnings to underreport.¹⁴

Studies of Program Participation and Transfer Income

The previous discussion focused on income reporting. There are also several older studies (1940 to 1984) of transfer program reporting in surveys. These are not “complete” design studies, in that they typically focus on a sample of recipients and examine whether they report benefits. Complete designs also would look at nonrecipients to see if they falsely report receipt. More recent studies do the latter. Most, but not all, of these studies find fairly substantial underreporting of transfer program receipt.

Marquis and Moore (1990), using two waves of the 1994 SIPP panel, did a comprehensive study of the accuracy of reporting of transfer program participation. They found evidence of substantial underreporting of program participation among true program participants, on the order of 50 percent for Workers’ Compensation and AFDC, 39 percent for UI, and 23 percent for Food Stamps and SSI. Overall participation rates for transfer programs, however, were quite close to what would be expected from administrative controls.

Subsequent work by Moore, Marquis, and Bogen (1996) on a sample of households from Milwaukee found smaller underreporting among true recipients. They concluded that most error, when it exists, is due to participants’ failure to report the sources of income, rather than a failure to report all months of participation.

Bollinger and David (forthcoming) give a detailed examination to underreporting of Food Stamp participation in the 1984 SIPP panel. They find that the high rate of underreporting for Food Stamps arises in part from failure to locate the person legally certified within the household. About half of the underreports within a household were offset by an overreport from another household member. The net

¹⁴They also examine several measurement error assumptions that challenge standard practice in empirical economics.

effect was underreporting of Food Stamp receipt of 12 to 13 percent in the 1984 SIPP panel. Bollinger and David also document the important point that nonresponse and false answers are correlated across survey waves in the SIPP.

Finally, Yen and Nelson (1996) examine survey and administrative records from the state of Washington and find that 93 percent of the almost 49,000 person months are reported correctly, and net overreports roughly equal net underreports.

Assessment of Income and Transfer Program Reporting in National Surveys

Moore, Stinson, and Welniak (1997) conclude their survey of what is known about income measurement in surveys by stating that

wage and salary income response bias estimates from a wide variety of studies are generally small and without consistent sign, and indicators of unreliability (random error) are quite low. Bias estimates for transfer income amount reporting vary in magnitude but are generally negative, indicating underreporting, and random error also is an important problem. (p. 12)

They conclude that “the additional data continue to support [the] conclusion of very little bias in survey reports of wage and salary income, and little random error as well.” Furthermore, studies that match administrative records of transfer programs and survey data “suggest a general tendency for transfer program income to be at least modestly—and in some instances substantially—under reported” (p. 16).

Based on our review of available assessments of income and employment measurement in national surveys, we think the above quotation is still correct.

- The CPS, SIPP, NLS, and PSID surveys provide valuable information on the behavior of the low-income population (and many other issues). They have national samples, broad and fairly accurate measures of income, and their focus on families as the unit of analysis and their ease of access greatly enhance their value.
- The value of these data sets for evaluating welfare reform is severely limited, however. With the devolution of responsibility for TANF, the CPS and SIPP sampling frames and sample sizes mean that, at best, they can be only supplementary data sources for understanding the effects of

welfare reform at the state and local levels. The apparent decline in program coverage in the CPS is also worrisome.¹⁵

3. UNEMPLOYMENT INSURANCE WAGE RECORDS DATA TO MEASURE INCOME AND EMPLOYMENT FOR LOW-INCOME POPULATIONS

We now consider the evidence on using UI wage records to measure the income and employment status of low-income populations. UI wage records contain the earnings reported by employers (on a quarterly basis) to state UI agencies for each employee. As noted above, UI data are often linked to information on targeted samples, such as participants in evaluations of specific welfare or training programs. Thus, the populations for which UI wage data are used to measure their income and employment vary with the particular investigation being conducted. Below, we report on several of these studies, attempting to draw some general conclusions about the strengths and weaknesses of this data source.

3.1 Potential Strengths

Using UI wage records to measure income and employment has several advantages. The first is that wages reported to state UI programs are thought to include most of the wage earnings of individuals. By law, any employer paying \$1,500 in wages during a calendar quarter to one or more employees is subject to a state UI tax and, hence, must report quarterly what is paid to each employee, including regular earnings, overtime, and tips and bonuses. Agricultural employers must report earnings if they have either a quarterly payroll of at least \$20,000 or have hired ten or more employees in each of 20 or more weeks during the preceding calendar year. Employers of paid household help must report wages if

¹⁵We were not able to find a comparable study of trends in program participation for the SIPP. U.S. Department of Commerce, Bureau of the Census (1998) compiles summaries of an extensive, long-running research program on SIPP quality. It starts with an overview of SIPP design, and then describes sample selection, data collection, nonresponse and measurement error, data preparation, weighting, sampling error, and evaluation of estimates from the 1984 to 1993 panels, and the 1996 redesign of the SIPP.

they pay at least \$1,000 in cash wages during any quarter. In a study of the use of UI wage records to measure the postenrollment earnings of Job Training Partnership Act (JTPA) recipients, Baj, Trott, and Stevens (1991) state, “Virtually all jobs that most observers would consider appropriate targets for JTPA trainee placement are covered by the UI reporting system.” (We discuss this study further below.)

A second potential advantage of UI wage data is its presumed accuracy. Hill et al. (1999), for example, make the following, perhaps incorrect (see below) argument:

Employers are liable for taxes up to an earnings threshold. Because this threshold is quite low, there appears to be little incentive for employers to underreport earnings for most employees. Moreover, employers’ reports are used to determine unemployment benefits. Discrepancies between employer and employee reports upon application of unemployment benefits can result in employer sanctions.

Baj, Trott, and Stevens (1991) claim:

The accuracy of the reporting of money wages is unknown. However, relatively few corrections occur in the routine processing of individual unemployment insurance claims. In addition electronic payroll processing is increasing, electronic cross-matching capabilities are expanding, and new revenue quality control practices have been introduced. Thus, there is reason to think that the accuracy of UI data is higher than that of most self-reported sources of earnings information. Intentional underreporting of wages constitutes fraud, which is subject to sanctions. Unintentional misreporting is subject to penalty payments.

A third presumed advantage of using UI data to measure employment and wage income of individuals is their ready availability, at least for certain authorized studies, and the ability to link these data with information from other administrative or survey data sources. (To safeguard the confidentiality of this information, state UI authorities control access to UI wage records and the Social Security numbers [SSNs] necessary to link these data to other data sources for individuals.) UI wage records are commonly used in state-level evaluations of welfare reform and other social programs. As Baj, Trott, and Stevens (1991) conclude in their study of the feasibility of using UI wage data from different states to monitor the posttraining earnings outcomes of individuals who received training services in JTPA:

The findings from the first phase of this project indicate that JTPA and *any other program* [emphasis added] whose goal is to increase the employment and earnings of participants can use UI wage-record data with confidence. Obtaining post-program information from state UI systems is not only a viable option, it is far more cost-effective

than the current practice of gathering this information through contact with participants. Furthermore, UI data are of higher quality than corresponding survey-based information. (p. 30)

They found, for example, that the response rate to the survey was 70.2 percent for those who were employed at termination versus 49.6 percent for those who were not. Based on these results, they concluded that using UI wage data was preferred to obtaining data via surveys, especially given the cost of conducting surveys on this population.

To summarize, there are several potential strengths of using UI wage records to measure income and employment.

- UI data are available at the state level and can be matched to individuals in existing samples at relatively low cost (as long as SSNs are available).
- It is straightforward to do follow-up analyses on income and employment for workers who remain in the state.
- Data are timely, in that they become available with roughly a two-quarter lag.
- For most workers the reporting of wage and salary income appears to be accurate, but see the concerns in the following section.

3.2 Potential Weaknesses

There are two potentially serious weaknesses with relying on UI wage records to measure employment and income for low-income populations. The first arises because UI wage records do not cover all forms of employment. In particular, state UI systems typically do not cover self-employed persons, most independent contractors, military personnel, federal government workers, railroad employees, some part-time employees of nonprofit institutions, employees of religious orders, and some students employed by their schools. As such, wage earnings from these types of employment are not contained in state UI wage records.

It is not clear how important these exemptions are. In at least two places in the literature, an assertion is made that over 90 percent of workers in the U.S. economy are in jobs covered by the UI

system (Baj, Trott, and Stevens, 1991; Kornfeld and Bloom, 1999).¹⁶ As noted below, this statistic is challenged by the results of Blakemore et al. (1996) and Burgess, Blakemore, and Low (1998), but even if correct, it is not clear how comforting it should be if the topic of interest is low-wage labor markets. If, say, 8 percent of all jobs are missing from UI wage records, but all 8 percent are low-income workers (which in turn is a much larger fraction of all low-income workers), the usefulness of UI data in monitoring the effects of welfare reform would be severely eroded.

Blakemore et al. (1996) and Burgess, Blakemore, and Low (1998) report results of a fascinating study of 875 Illinois employers from 1987 that were subjected to detailed audits of their UI reports. As part of the data set, routine information such as the employment size of the firm, the statutory UI tax rate for each firm, one-digit SIC codes, and punctuality of UI reporting were compiled. Also included were unique audit information on unreported workers, underreported total and taxable wages, and UI taxes due on these unreported wages. The studies also merged information on the total number of independent contractors that each firm reported to the IRS. The data set does not attempt to identify employers who were part of the underground economy.

If the results for Illinois are projected nationally,¹⁷ employers failed to report the presence of 11.1 million UI-eligible workers and \$70.6 billion in wages to state UI agencies in 1987. This is 13.6 percent of all workers. Some of the undercoverage arose from failure to report casual or part-time workers, and failure to report tips, bonuses, and other types of irregular compensation. By far the largest problem (accounting for roughly 50 percent of the discrepancy), however, was with independent contractors. Issues surrounding independent contractors are among the most vexing in tax administration and labor law. In brief (and at the risk of oversimplification), in tax law there is a somewhat subjective, 20-part test

¹⁶Despite our efforts, we have not found documentation for this statistic.

¹⁷Clearly, strong assumptions are needed to make this projection, but the size and industrial composition of the Illinois sample are not sharply different from national statistics. The Illinois UI system is typical of what is observed nationally, and, if anything, Midwestern states tend to have lower rates of income and payroll tax noncompliance than other states.

to define a worker as a regular employee or independent contractor. Elements of the test include (from IRS Publication 15A) whether the business has “behavioral control” of the worker (does the business give instructions and train the worker); financial control (can the worker make a profit or loss, does the worker have unreimbursed business expenses, does the worker make services available to a broad market); and type of relationship (does the job have benefits, is it permanent, are the tasks a key aspect of the regular business of the company). If a worker is treated as an independent contractor, an employer does not have to withhold income taxes, withhold and pay Social Security and Medicare taxes, or pay UI taxes.

It is not clear if the issues raised in the Illinois UI audits are strictly associated with independent contractors (in the technical sense) or more broadly with flexible staffing arrangements. Houseman (1999) provides a nice introduction to issues associated with flexible staffing arrangements. She reports data from the February 1997 CPS Supplement on Contingent and Alternative Work Arrangements that 6.7 percent of workers were “independent contractors,” 1.0 percent were “agency temporaries,” 1.6 percent were “on-call or day laborers,” 0.6 percent were “contract company workers,” and 2.6 percent were “other direct-hire temporaries.” These categories constitute 12.5 percent of the workforce. The use of flexible staffing arrangements appears to have been growing sharply over time, but detailed information on this growth is not available. Houseman (1999) reports that the IRS estimates losing billions in tax revenues each year due to misclassification of employees.

Houseman (1999) also reports information on the incomes of “flexible workers” drawn from the February 1995 CPS Supplement on Contingent and Alternative work, matched to the March 1995 CPS. Some 7.5 percent of “regular” employees had incomes below 125 percent of poverty. The corresponding figures for agency temporaries was 21.7 percent; 16.2 percent for on-call or day laborers; 10.8 for independent contractors; 11.5 for contract company workers; and 15.1 percent for other short-term direct hires. Consequently, a failure of UI data to fully capture workers in flexible staffing arrangements could pose a major problem for studies that rely exclusively on UI data to measure the income and employment of low-income workers.

In many industries, employers have considerable flexibility in designating the status of workers. At least in the Illinois audit study, employers aggressively overused the independent contractor designation. In all, 45 percent of all employers make some underreporting error. This includes almost 500,000 cases in which workers were erroneously excluded, resulting in \$2.6 billion in wages being underreported. Smaller firms were estimated to underreport 14 percent of their taxable wages and 56 percent of their UI-covered workforce. In statistical models, the percentage of workers on the payroll who are independent contractors and the turnover of the firms' workers are two key explanatory variables. The effective tax rate, though related to turnover, also appears to be positively associated with compliance. The characteristics of firms that make errors on UI reports would appear to be positively correlated with the type of employers who disproportionately hire workers with low levels of human capital.

Hence, we view the Blakemore et al. (1996) and Burgess, Blakemore, and Low (1998) studies as raising a very serious concern about the coverage of UI data, and hence its suitability as the *exclusive* source of data with which to evaluate welfare reform. In our conclusions, we recommend that at least one additional study be conducted along the lines of the one in Illinois to assess UI coverage. It is our impression, based on casual, anecdotal evidence, that the use of independent contractors has increased fairly substantially over time, and hence the work based on 1987 Illinois data may understate the problem.

The second potentially major weakness with using UI data for evaluating welfare reform is that it contains limited accompanying demographic information on individuals, and, more important, may not allow one to form an accurate measure of family income. In assessing the impacts of welfare reform, many argue that it is important to assess how these changes affect the well-being of children and the families in which they reside. As such, families constitute the natural "unit of analysis" for such assessments, and family income is often used as an indicator of this unit's well-being.

The potential problem of relying on earnings data from UI wage records when the objective is to assess the level of family resources in studying the impact of welfare reform has recently been highlighted by Rolston (1999). Based on past research, Rolston notes that changes in individual income account for

only 40 to 50 percent of exits from welfare. Thus, to have a complete picture of the effects of welfare reform, analysts need information on other economic and demographic changes occurring in the family. Given this, the problem is clear. Income as reported through UI records fails to include nonemployment income and income of partners that become available to a family and whose availability allows the family to not receive cash assistance or in some cases renders them ineligible for it.

Calculations from Meyer and Cancian (1998) suggest that the concern raised by Rolston (1999) is economically important. Recall that Meyer and Cancian found, for example, that 5 years after leaving welfare 64.2 percent of the women still had incomes below the poverty line, while, when considering the broader family unit, only 40.5 percent had income below the poverty line. In a related calculation, however, Primus et al. (1999) perform an analysis that shows “for most single-mother families, including the income of unrelated male individuals does not materially change the picture drawn of a decline in overall disposable income between 1995 and 1997.” More needs to be learned about the importance of the issue raised by Rolston in assessing the level and trend in family well-being following welfare reform.

To summarize, there are two potential weaknesses of using UI wage records to measure income and employment.

- UI data do not cover all workers, including the self-employed, military, federal employees, independent contractors, and other employment arrangements. Some evidence suggests that the gaps in coverage may be significant.
- UI data follow individuals, so one cannot obtain information on incomes of other family members, at least without SSNs of other household members. UI data also provide limited demographic and background information on workers.

3.3 Direct Comparisons of UI Wage (and Employment) Data with on Income and Employment Measures from Surveys

In this section, we review two sets of studies that directly compare income and employment measurement across several data sources for the same individual and/or family. We first consider the results of a comparison of measures of income and employment gathered from UI records in 11 states and from a survey for a sample of 42,564 adults who left JTPA programs during the 1986 program year. The

findings from this study are described in Baj, Trott and Stevens (1991) and Baj, Fahey, and Trott (1992).¹⁸ Of those surveyed, 27,721 responded to all three of the questions that were mandatory for “terminees” of JTPA-sponsored programs, giving an overall response rate of 65.1 percent. The investigators had access to separate data files containing UI wage records for the full sample of terminees, where the latter information was drawn from the UI systems for the 11 Midwestern states included in their study. Baj, Trott, and Stevens drew the following conclusions about estimating the postenrollment incomes of these JTPA terminees with these two alternative data sources:

There are two major conclusions to be drawn from these analyses. First, there is ample evidence to suggest that the post-program survey data is substantially affected by the presence of non-response bias. While this conclusion is based largely on the examination of post-program employment experiences, it is suspected that the same conclusion would hold if the focus was on post-program earnings. The second conclusion is that the major source of this bias, i.e., the different post-program employment experiences of respondents and non-respondents who were employed at termination, is not addressed through current non-response adjustment procedures. The implication of these findings is that the estimates of post-program performance based on the information gathered through the post-program survey are not a true reflection of the actual post-program experiences of all JTPA terminees. (p. 35)

The survey they examined was not constructed in a way that allows comparisons of earnings reports. Instead, the presence of employment in a given quarter was compared across the survey and UI data. To do this, they sharply restrict the sample to people leaving Title II-A (JTPA) one week prior to the week containing the starting date of a fiscal quarter.¹⁹ For data reasons, three states also were dropped from the sample. This left 1,285 participants, 863 of whom responded to the survey. Even with these sample restrictions, employment comparisons are not completely straightforward because UI earnings are reported for the quarter in which they are paid, not the quarter in which they are earned. With these issues in mind, Table 4 shows the results of the comparisons.

¹⁸Stevens, Chen, and Shi (1994) conducted a similar study focusing on Maryland.

¹⁹Title II-A is the nation’s employment and training program for low-income adults and out-of-school youths with significant barriers to employment.

TABLE 4
Comparison of the Employment Status Results for the 13-Week Program Survey
and UI Wage Record Data, Title II-A Adult Survey Population

Postprogram Survey Status	First Quarter UI Status		Total
	Employed	Unemployed	
Employed	628 (72.8%)	114 (13.2%)	742 (86.0%)
Unemployed	44 (5.1%)	77 (8.9%)	121 (14.0%)
Total	672 (77.9%)	191 (22.1%)	863 (100%)

Source: Baj, Trott, and Stevens (1991), p. 39.

The diagonal elements show that 81.7 percent of the UI-survey observations are in agreement on employment status. The lower off-diagonal element indicates that 5.1 percent of the matched sample report that they were unemployed during the quarter, yet they had UI earnings. One might think welfare recipients would be reluctant to report earnings, but they were only slightly (5.4 percent) more likely to not report earnings (when they had positive UI earnings) than nonrecipients (4.4 percent). There are two potential explanations for this result. First, respondents may have earned the UI wages reported for the quarter during the previous quarter and subsequently lost their jobs. Second, respondents may have provided inaccurate reports. Given that many of these 44 cases were employed at the time they left JTPA, Baj, Trott, and Stevens suggest that the second explanation is more likely than the first.

The upper off-diagonal element shows that 13.2 percent of the sample report being employed yet without UI wages.²⁰ Again, it is possible that the timing of UI wage reports can partially account for this discrepancy, though most of these people were employed at the time they left JTPA, so this again is an unlikely explanation. Instead, it is likely that some of these people were employed out of state, and that others had jobs that were not covered by UI.²¹

Baj, Fahey, and Trott (1992) update the Baj, Trott, and Stevens (1991) calculations and provide more detail on the potential sources of discrepancy between UI data and the survey that was administered. In 1987, 11.3 percent of the sample report being unemployed for the quarter but have UI data (the corresponding figure from the earlier study was 5.1 percent), and 9.1 percent have no UI record but report they are employed (the corresponding figure from the earlier study was 13.2 percent). Baj, Fahey, and Trott discuss three possible reasons to explain cases that claim to be employed but show no UI record.²²

²⁰Though this figure appears comparable to Blakemore et al. (1996), it actually suggests a much smaller gap in coverage. This study found a 13.6 percent gap for the total workforce, while the Baj, Trott, and Stevens (1991) found that the corresponding gap is 13.2 percent for a JTPA sample.

²¹Tabulations from the Wisconsin Department of Workforce Development (1999) suggest that as many as 16 percent (60/375) of a small sample of recent welfare recipients have missing employment episodes in UI data.

²²They do not discuss independent contractor issues that are the focus of Blakemore et al. (1996). Instead, these would be grouped into the last category.

The respondent may have been employed out of state, employed in the quarter but have wages that were not paid until the next quarter, or were employed in a job not covered by UI or where the employer failed to report UI wages.

To look at these factors, the authors use data from the Illinois JTPA management information system, which gives detailed information on the employment status at termination of the program and compares that to UI status at termination. The analysis focuses on 3,387 cases (13.1 percent of the sample) that reported that JTPA participants were employed at termination, but for which there was no UI record for the termination quarter. Table 5 suggests some explanations for the mismatches (at the termination quarter). The table shows that out-of-state employment accounts for 15.3 percent of the discrepancies (row 1). Identifiable employment in uncovered (self-employed and federal appointments) sectors accounts for 6.6 percent of the discrepancy (rows 2 and 3). The next three rows of the table—within program, first quarter, and second quarter UI entries—is supposed to reflect timing differences in the data. These collectively account for 23.1 percent of the discrepancy (rows 4, 5, and 6). Another 15.9 percent of the discrepancies seem to result from name mismatches between employers that could be fairly easily reconciled. This still leaves 39.1 percent of the remaining sample unexplained. Of this group of 1,325 participants, there were 1,108 different employers. The potential explanations for the discrepancy that Baj, Fahey, and Trott (1992) offer include errors in reporting the Social Security number on the JTPA or UI data systems, an employer's neglect of UI reporting requirements, and reporting errors by JTPA operators.

Baj, Trott, and Stevens (1991) and Baj, Fahey, and Trott (1992) examine the existence of employment in survey and UI data, but do not provide comparisons of earnings because their survey did not elicit information on earnings. Kornfeld and Bloom (1999) look at both employment and earnings. They describe their study as attempting “to determine whether wage records reported by employers to state unemployment insurance agencies provide a valid alternative to more costly retrospective sample surveys of individuals as the basis for measuring the impacts of employment and training programs for

TABLE 5
Causes for Mismatches: Participants Employed at Termination with No UI Record
for the Quarter of Termination, Illinois Program Year 1987 Title II-A Adult Population

Reason for Mismatch	# of Cases	Percent
Employed out of state	517	15.3
Self-employed	51	1.5
Federal employment	172	5.1
Within program UI record	81	2.4
1st quarter UI record	608	18.0
2nd quarter UI record	93	2.7
No related UI record	1,865	55.0
No UI record	1,325	39.1
Mismatched employers	540	15.9
Total	3,387	100.0

Source: Baj, Fahey, and Trott (1992), p. 142.

low-income persons” (p. 168). The Kornfeld and Bloom study is based on data covering 12,318 people from 12 sites around the country in which an experimental evaluation of JTPA training programs was conducted. For each site, they had access to data from both UI wage records and follow-up surveys of experimental (received JTPA services) and control (did not receive JTPA services) group members. In their analysis, they dropped observations with missing or imputed data, but included observations where earnings were recorded as zeros in the follow-up surveys.

Kornfeld and Bloom (1999) assess how UI and survey data differ in estimating the *levels* of earnings and the *differences* in mean earnings and employment rates between experimental and control group members. Although the primary objective of the Kornfeld and Bloom study is to assess how the estimated impacts of JTPA services on income and employment status vary by data source—they find virtually no difference in the estimates of impact by data source—we shall focus on what they found with respect to differences in levels of earnings across the two sources of income and employment data available to them.

Table 6, drawn from their study, shows that employment rates calculated from the two data sources are quite close. The discrepancies between employment data derived from their survey versus from UI records range anywhere from employment being 1 percent lower in surveys to being 11 percent more. At the same time, Kornfeld and Bloom find that the discrepancies in the *level* of earnings for JTPA participants are much greater. In particular, they consistently find that the level of earnings from survey data is higher than in UI data. The nature of this discrepancy in earnings measures is different from the one raised in Rolston (1999). Rolston is concerned that using UI wage data to measure the earnings of welfare leavers tends to be biased because such data do not include the income of other family members. Rolston argues that this lack of inclusion of earnings of other family members is important given evidence that suggests that many exits from welfare are coincident with changes in family structure. The comparison in Table 6 from Kornfeld and Bloom (1999) focuses only on earnings reports for individuals. It documents systematic discrepancies of UI and survey data, where income reported by UI data is always

TABLE 6
Comparison of Quarterly Earnings and Employment Rates from UI and Survey Data

	Treatment Earnings (\$)	Control Earnings (\$)	Treatment Employment Rate	Control Employment Rate
Adult Women (4,943; 18,275; 8,916)				
Survey data	\$1,294	\$1,141	59.2%	54.5%
UI data	\$1,048	\$922	57.6%	54.1%
Ratio (Survey/UI)	1.23	1.24	1.03	1.01
Adult Men (3,651; 13,329; 6,482)				
Survey data	\$1,917	\$1,824	65.8%	63.5%
UI data	\$1,456	\$1,398	61.7%	60.7%
Ratio (Survey/UI)	1.32	1.30	1.07	1.05
Female Youths (2,113; 9,452; 4,316)				
Survey data	\$951	\$949	51.3%	50.6%
UI data	\$701	\$700	50.6%	51.2%
Ratio (Survey/UI)	1.36	1.36	1.01	0.99
Male Youths without a Prior Arrest (1,225; 5,009; 2,442)				
Survey data	\$1,556	\$1,655	65.5%	69.3%
UI data	\$1,015	\$1,103	61.3%	63.2%
Ratio (Survey/UI)	1.53	1.50	1.07	1.10
Male Youths with a Prior Arrest (386; 1,646; 705)				
Survey data	\$1,282	\$1,531	58.0%	61.3%
UI data	\$759	\$760	52.8%	55.0%
Ratio (Survey/UI)	1.69	2.01	1.10	1.11

Source: Kornfeld and Bloom (1999), Tables 1 and 2.

Note: Numbers after each panel heading reflect the number of persons represented (e.g., 4,943 adult women) with the number of person-quarters in the treatment and control groups.

substantially lower (in one case, by half) than that reported in survey data. Because the employment rates are comparable, Kornfeld and Bloom conclude that the earnings differences must reflect either differences in hours of work for JTPA participants who are recorded as being employed in a quarter, differences in the rate of pay recorded for this work, or both.

Kornfeld and Bloom (1999) also condition on whether a JTPA participant was receiving AFDC benefits during a particular quarter and find that, while the level of earnings is lower, the discrepancy between survey and UI data is strikingly similar for AFDC recipients and nonrecipients. Survey earnings reports for adult women and female youths are 24 to 34 percent higher than reported UI earnings levels. There was also wide variation across JTPA sites in the size of earnings discrepancies between survey and UI data, but the survey always yielded larger numbers than did the UI data. The “ratio range” was 1.15 to 1.40 for adult women, 1.16 to 1.72 for adult men, 1.16 to 1.76 for female youths, and even larger for male youths. Whatever the mechanism generating these discrepancies, it exists across all 12 geographically diverse JTPA sites.

The dispersion of earnings discrepancies is very large, so the means mask large variation across earnings reports. We do not know, of course, which measure of earnings more closely reflects the truth. If survey data tend to be more accurate, however, the discrepancies shown in Table 7 would give one pause in using UI data to assess the economic well-being of families following welfare reform. Table 7 shows that more than 10 percent of women and 20 percent of men have discrepancies that exceed \$1,000 in a quarter.²³

Kornfeld and Bloom (1999) also examine those JTPA participants for whom they found positive earnings in one data source but not the other. “When only the survey reported employment (and UI data presumably missed it), mean earnings were more than twice what they were when only UI data reported employment (and the surveys presumably missed it). This suggests that surveys are more likely to miss

²³All zeros in the table correspond to people without earnings in both data sets. No observations with positive earnings agreed exactly.

TABLE 7
Distribution of Mean Individual-Level Differences between Survey and UI-Reported
Quarterly Earnings

Mean Survey - Mean UI	Adult Women	Adult Men	Female Youth	Male Youth No Arrest	Male Youth with Arrest
\$2,001 +	3.5	9.9	2.0	8.7	9.3
1001-2000	7.7	11.1	7.7	15.0	16.1
601-1000	7.9	9.4	9.1	12.4	13.2
401-600	6.8	5.9	8.3	8.2	7.3
201-400	10.4	8.6	13.0	11.9	9.6
1-200	17.3	12.7	20.6	13.2	14.8
0	14.3	8.1	10.0	3.8	5.7
-\$1- -\$200	16.2	13.2	17.3	10.8	11.1
-201- -400	6.2	6.2	5.3	5.6	6.0
-401- -600	3.3	4.2	3.3	3.4	1.6
-601- -1000	3.4	4.6	2.0	4.5	3.4
-1001- -2000	2.3	4.1	1.1	1.7	1.6
-\$2,001-	0.9	2.0	0.2	0.8	0.5
Mean diff (\$)	228	451	256	547	605

Source: Table 5 from Kornfeld and Bloom (1999).

‘low earnings’ quarters, perhaps because respondents forget about minor, or short-term, jobs. In contrast, UI data appear more likely to miss ‘average earnings’ quarters—where mean earnings are similar to when both data sources report employment. This might be due to random errors in matching UI wage records, out-of-state jobs, jobs that are not covered by UI, and/or earnings that are ‘off the books’” (p. 184).

The above-noted discrepancies could arise between the data sources because some jobs are uncovered, some jobs may be located out of state, some payments may go unreported because of unintentional or intentional noncompliance, or Social Security numbers may be misreported. To provide further insight, Kornfeld and Bloom compare the earnings reports that employers make about their employees to state UI systems with those they make to the IRS. Employers have an incentive to underreport earnings to the UI system (and hence avoid paying UI taxes), but they have no incentive to conceal earnings when reporting to the IRS, since wages are a business expense that will lower tax payments. The sample for doing this comparison is smaller than the previous samples since each observation needs to be there for four consecutive quarters, corresponding to the calendar year. The ratio of mean IRS earnings to mean UI earnings ranged from 1.14 for adult women to 1.25 for male youths, so UI wage records clearly are missing earnings from some jobs.

Kornfeld and Bloom draw the following conclusions from their investigation.²⁴ Approximately one-half of the survey-UI earnings difference reflects earnings that are missing from UI wage records (by making use of the IRS data). Out-of-state jobs do not explain why UI wage records reported lower earnings than sample surveys. Uncovered jobs account for only a small part of the survey-UI earnings difference. There is little evidence consistent with recall bias in the survey data. There is no evidence that

²⁴A third comparative study assessing differences in income and employment across data sources was conducted by the Rockefeller Institute of Government (1999). This study summarizes (in its Table 2) six studies that compare UI data and survey data. The studies include Kornfeld and Bloom (1999) and five that we do not review in this paper. The results from the Rockefeller Institute study differ somewhat from the Kornfeld and Bloom results. One of two other studies finds that UI earnings are lower than survey data (though one found them nearly identical), like Kornfeld and Bloom. Several other studies suggested that employment rates from surveys were significantly higher (on the order of 20 percent) than employment rates from UI data, unlike Kornfeld and Bloom. We have not estimated the quality of these other studies.

large survey discrepancies result from survey reports of “unusually good” jobs or weird reports of industry of employment. And survey discrepancies also do not appear to be driven by overtime or odd pay periods.

From direct comparisons of the data sources used to measure income and employment status in the studies reviewed above, we draw the following tentative conclusions about using survey versus UI data.

- Earnings in UI data appear to be generally lower than earnings reported in survey data. The UI data may miss earnings from second or casual jobs. At the same time, surveys may *overstate* earnings. Smith (1997) provides a thorough comparison of a data set of JTPA-eligible nonparticipants at four of the 16 JTPA training centers with data from SIPP. He provides evidence that his JTPA survey may be biased upward, due to nonresponse bias (lower earners are not covered in the survey) and upward-biased measures of overtime and usual hours in the survey.
- Employment rates derived from UI data are comparable to lower than those that result from survey data. We expect UI-based employment rates to be lower because of coverage problems with flexible workers/independent contractors. Surveys also suffer from nonresponse, however, so undercounts in both data sources may be comparable, making the UI-based rates similar to survey-based rates.

3.4 Overall Assessment of Using UI Administrative Data to Measure Income and Employment

Our review of the literature has pointed to three critical concerns that arise with using UI data to measure the earnings and employment of low-income and welfare-eligible populations.

- First, earnings are available only for individuals, while changes in family composition upon exit from welfare have been shown to have a large bearing on economic well-being. UI data do not allow us to track these changes.
- Second, there appears to be a substantial problem with some workers being classified as independent contractors and hence not entering the UI system. Overall, gaps in coverage appear to be at least 13 percent and may be significantly higher.
- Third, even when wages are reported, they appear to be understated by at least 11 to 14 percent (based on the Kornfeld and Bloom comparisons with IRS data) and perhaps more than twice that (based on their comparisons with survey data). Smith (1997) shows, however, that survey responses can also be biased upward.

4. DATA FROM FEDERAL (AND STATE) INCOME TAX RETURNS

Though not widely used in past evaluations, wage and salary data from federal and state income tax returns represent an alternative to UI data for measuring the income and employment of low-income populations. Below we outline the potential strengths and weaknesses of this data source and briefly summarize a recent comparison of UI wage and tax return data for a disadvantaged population drawn from the AFDC caseload in California.

4.1 Potential Strengths

Compared with using surveys or UI records, tax returns have at least two potential advantages for measuring income and employment.

- The data from tax returns are accurate. Taxpayers provide information under the threat of audit and there is third-party information reporting, so employers as well as recipients are reporting wage and salary information.
- The definition of income that is reported is broader than that provided by unemployment insurance data, including especially self-employment income and, in cases where a person is married and filing a joint return, spousal income.²⁵

4.2 Potential Weaknesses

Several potential weaknesses are associated with using data from tax returns to measure income and employment. We summarize these below. Note that some of the weaknesses apply to the general population, while others are more relevant for low-income populations. First, the access by researchers to tax return data is extremely limited and constrained because of Section 6103 of the Internal Revenue Code. Section 6103 explicitly states that tax data cannot be released, except to organizations specifically designated in Section 6103(j). The exceptions are the Department of Commerce, but only as it relates to the Census and National Income Accounts, the Federal Trade Commission, the Department of Treasury,

²⁵It also includes interest and dividend income, farm income, capital gains and losses, and gambling winnings, as well as indicate recipients of government transfers and Social Security benefits.

and the Department of Agriculture (for conducting the Census of Agriculture). Penalties for unauthorized disclosure are severe, including jail terms of up to 5 years.

Second, tax return data also contain only limited information on demographic characteristics of taxpayers. For example, the tax system does not collect information on the race or education of tax filers.

Third, it is also the case that tax-filing units differ from both families and individuals. Married couples can file either a joint return or separate returns (as “married filing separately”). Cohabiting couples, even if fully sharing resources, will file separate returns as individuals or heads of household (generally meaning the filer is a single parent with dependents). In general we think families pool resources, so families are the best unit of analysis for assessing economic well-being. Hence, case units probably are the most useful unit of analysis, followed by tax returns and then unemployment insurance data.

Fourth, there also are differences between tax return data and other data sources in the frequency of reporting. Unemployment insurance wages are reported quarterly. Transfer program information is reported monthly. Tax returns are filed annually. Since shorter periods can be aggregated into longer ones and since major changes in family composition can occur over time, the annual frequency of tax reporting is less appealing than monthly or quarterly reporting in other data sets. To the extent that family structure changes over these intervals, problems may arise when trying to link different data sets to assess well-being.

A fifth concern relates to the incidence and accuracy of tax filing by individuals and households, especially among low-income populations. This concern takes two forms: (1) do people file any tax return

and (2) if they file, do they report all sources of income to the IRS (or state taxing authorities)?²⁶ We consider each in turn.

If large fractions of low-income taxpayers do not file tax returns, then tax return data have very limited value. Unfortunately, there is not a lot of information on the filing propensities of people with low income. Information from the early 1990s (Scholz, 1994) suggested that 14 to 20 percent of those entitled to the EITC at the time failed to receive it, meaning that they failed to file tax returns.²⁷ Below, we discuss one recent study on the tax filing propensities of a low-income population that sheds some preliminary light on this issue.

Among filing units, it is also possible that their members do not report all of their sources of income on their tax returns. For example, individuals may fail to file income received as independent contractors. Firms or individuals who make use of independent contractors are obligated to report payments to such contractors to the IRS, but failure to do this generally is difficult to detect. Again, we know very little about the incidence of underreporting of various income sources for low-income populations.

To summarize, there are several potential weaknesses of using tax return data to measure income and employment.

- It is difficult to gain access to tax returns.
- They have limited information on demographic and other characteristics.
- Some low-income workers may not file, despite being eligible for the EITC, or may not report all their income.

²⁶If one is just interested in enumerating the population (as opposed to knowing incomes associated with families and individuals within the population), IRS data appear to be comprehensive. Sailer and Weber (1998–1999) report the IRS population count is 95.4 percent of the census population count. The consistency is fairly good across gender, age, and state. Unfortunately, for many of the people enumerated, the IRS does not know anything about them other than that they exist.

²⁷Cilke (1998) uses a CPS-IRS exact match file to examine characteristics of people who are not required to file tax returns and actually did not file tax returns. The entire paper is presented as proportions, however, so it does not provide information on the absolute number of low-income families with earnings who fail to file.

4.3 Comparison of Income Reporting from UI Wage and IRS Tax Filings Data for a Low-Income Population

In a recent study of the EITC for a sample of assistance units on the California caseload, Hill et al. (1999) compared UI wage data with linked data from the sample members' IRS tax returns. The study used data from the California Work Pays Demonstration Project (CWDP), which was conducted in four counties (Alameda, Los Angeles, San Bernardino, and San Joaquin) starting in 1992. The data consisted of two sets of assistance units drawn from the caseloads in these counties. One set, the sample used in Table 8, consisted of a random sample drawn from caseload at a particular date in 1992. Although this sample is representative of the caseload at that time, recall that the study by Bane and Ellwood (1983) showed that random samples from the existing caseload of AFDC are disproportionately made up of assistance units that are "welfare dependent."

The second set of assistance units, the sample used for Table 9, is a random sample of new entrants to the caseload in 1993. Bane and Ellwood (1983) and others have found that a significant proportion of new entrants remain on welfare for only a relatively short period.²⁸ Furthermore, Gritz and MaCurdy (1992) find that most new entrants exit from AFDC to employment. We also break both samples up into female-headed households (AFDC-FG cases) and two-parent households (AFDC-U). We report on annual earnings information for the year after the samples were drawn, i.e., 1993 for the random sample of the caseload and 1994 for the new entrants sample.²⁹

The first two rows of each panel of each table give estimates of the employment rates of each sample of AFDC recipients. As expected, employment rates of the point-in-time caseload (Table 8) are

²⁸For example, Bane and Ellwood (1983) estimate that 65 percent of new entrants leave the caseload in 2 years or less.

²⁹Through an interagency agreement between the California Department of Social Services and the state's taxing authority, the Franchise Tax Board (FTB), UI wages and wages and adjusted gross income from tax returns were merged by the FTB. The researchers were able to specify computer runs on these merged files. Assistance units in the study could, and did, leave AFDC after they were enrolled in this study. Nonetheless, wage and income data from UI records and tax returns were available for all of the original assistance units in the CWDP study.

TABLE 8
Random Sample from Caseload in 1992, Information for Tax Year 1993

	Those Filing Tax Returns	Full Sample
AFDC-FG Cases		
% of households with UI earnings		26
% of households that filed tax returns		22
Average UI earnings of adults in household	\$4,514	\$1,242
Average adjusted gross earnings on tax returns	\$10,589	\$2,378
Average wage & salary earnings on tax returns (Line 7)	\$9,748	\$2,189
Average income reported to AFDC	\$1,222	\$360
% of households with no UI earnings, but filed tax return		5.89
% of households with UI earnings, but filed no tax return		11.41
% of households for which AGI < UI wages		12.61
% of households for which AGI = UI wages		78.59
% of households for which AGI > UI wages		8.80
% of households for which AGI < UI wages, for UI wages > 0		3.39
% of households for which AGI > UI wages, for AGI > 0		40.47
<i>Self-Employment Income Reported on Tax Returns</i>		
Fraction of filers reporting any	0.06	
Average amount reported	\$357	
AFDC-U Cases		
% of households with UI earnings		31
% of households that filed tax returns		31
Average UI earnings of adults in household	\$5,223	\$1,792
Average adjusted gross earnings on tax returns	\$8,482	\$2,595
Average wage & salary earnings on tax returns (Line 7)	\$7,554	\$2,311
Average income reported to AFDC	\$2,513	\$894
% of households with no UI earnings, but filed tax return		7.07
% of households with UI earnings, but filed no tax return		8.21
% of households for which AGI < UI wages		9.26
% of households for which AGI = UI wages		78.39
% of households for which AGI > UI wages		12.05
% of households for which AGI < UI wages, for UI wages > 0		3.97
% of households for which AGI > UI wages, for AGI > 0		40.53
<i>Self-Employment Income Reported on Tax Returns</i>		
Fraction of filers reporting any	0.12	
Average amount reported	\$562	

Source: Hill et al. (1999).

TABLE 9
Random Sample of New Entrants to AFDC Caseload in 1993, Information for Tax Year 1994

	Those Filing Tax Returns	Full Sample
AFDC-FG Cases		
% of households with UI earnings		37
% of households that filed tax returns		33
Average UI earnings of adults in household	\$6,769	\$2,868
Average adjusted gross earnings on tax returns	\$13,185	\$4,342
Average wage & salary earnings on tax returns (Line 7)	\$12,575	\$4,141
Average income reported to AFDC	\$1,625	\$709
% of households with no UI earnings, but filed tax return		8.34
% of households with UI earnings, but filed no tax return		13.55
% of households for which AGI < UI wages		15.69
% of households for which AGI = UI wages		71.31
% of households for which AGI > UI wages		13.00
% of households for which AGI < UI wages, for UI wages > 0		4.21
% of households for which AGI > UI wages, for AGI > 0		39.88
<i>Self-Employment Income Reported on Tax Returns</i>		
Fraction of filers reporting any	0.04	
Average amount reported	\$95	
AFDC-U Cases		
% of households with UI earnings		48
% of households that filed tax returns		49
Average UI earnings of adults in household	\$8,516	\$5,138
Average adjusted gross earnings on tax returns	\$12,970	\$6,360
Average wage & salary earnings on tax returns (Line 7)	\$11,421	\$5,601
Average income reported to AFDC	\$3,264	\$1,831
% of households with no UI earnings, but filed tax return		10.45
% of households with UI earnings, but filed no tax return		7.94
% of households for which AGI < UI wages		11.77
% of households for which AGI = UI wages		64.71
% of households for which AGI > UI wages		23.51
% of households for which AGI < UI wages, for UI wages > 0		6.12
% of households for which AGI > UI wages, for AGI > 0		46.83
<i>Self-Employment Income Reported on Tax Returns</i>		
Fraction of filers reporting any	0.11	
Average amount reported	\$512	

Source: Hill et al. (1999).

lower than the sample of new entrants (Table 9). Employment rates of one-parent cases (AFDC-FG) are lower than the employment rates of two-parent cases (AFDC-U). What is striking and not necessarily expected, however, is that the implied employment rates using UI data and using tax return data are nearly identical. From Table 8, employment rates of the point-in-time AFDC-FG caseload were 26 percent using UI data and 22 percent using tax return data. The corresponding rates for AFDC-U cases were 31 percent for both data sources. Employment rates were 37 percent using UI data for the new entrant sample, and 33 percent using tax returns. Employment rates were 48 percent using UI data for the AFDC-U new entrants and 49 percent using tax returns.

Tax return data and UI data would give similar perspectives about employment patterns of the four-county California sample, but it is clear that each sample covers workers that the other misses. For example, in the top panel of Table 8 (AFDC-FG cases from the point-in-time sample), roughly one-quarter of people (5.89/22) who filed tax returns had no corresponding UI record.³⁰ Over 40 percent (11.41/26) of those with positive UI earnings did not file taxes.³¹ Of those with both UI and tax return earnings, over 40 percent reported more earnings on tax returns than would be expected based on UI data. Similar figures apply to each other group, though for AFDC-U cases, only about 20 percent of the cases with UI earnings do not file tax returns.

The fact that across all four groups (two samples, and AFDC-FG and AFDG-U cases) tax return income exceeded UI income in at least 40 percent of the cases with positive earnings from both sources is consistent with households from this welfare-based population having earnings that are not from covered employment. The fact does not seem to be explained by people leaving welfare (through changes in family structure). Among AFDC-FG cases, only 1 to 13 percent of these households had no months on

³⁰We took great care in the analysis to make sure that the comparison samples did not have changes in marital status and had a full four quarters of UI data (including zero quarters).

³¹Households with low earnings are not obligated to file tax returns. A married couple, for example, is not required to file if their income is below the standard deduction and two exemptions (\$12,200 in 1997), regardless of how many children they have. Hill et al. (1999) also show that most of these nonfilers had very low levels of UI earnings (\$2,500 or less in annual covered earnings).

AFDC during the tax reference year, and between 56 and 83 percent were on welfare for between 9 and 12 months during that year. There is also little evidence that self-employment income plays an important role in earnings differences between tax return and UI income.

Based on comparisons between UI and tax return data, we offer several tentative conclusions.

- Tax return and UI data appear to give very similar information on employment rates of the four-county California caseload. There are good reasons, however, to think that both data sources will lead to underestimates. UI data will miss independent contractors and possibly other “flexible workers.” Tax return data will miss families who do not file tax returns.
- The two data sources appear highly *complementary*. Each appears to capture a significant number of families that the other misses. Using them together, therefore, should result in more accurate measures of the employment experiences of the caseload than using either separately.
- Tax return data have a broader definition of income and, if the household unit is married, will cover both spouses and hence is likely to offer more accurate income information.
- UI data are much easier to access than are tax returns.

5. RECOMMENDATIONS

Taking into account all features of a data source—including its accuracy, its cost, and its ease of access—no single source can be declared “preferred.” The inability to find a preferred data source is inevitable given the differences in the desired uses of data, the constraints imposed by budgets for data collection, and the access limitations to data. The fact that UI wage data are inexpensive, timely to obtain, and available at the state level, for example, implies that it will continue to be a focal data set for state-level evaluations of welfare reform. Our review, though, raises a number of serious questions about UI data. In the remainder of this paper, we highlight selected issues that we think need further attention in the hopes of encouraging future research on at least some of them.

Certain questions related to welfare reform can only be answered with nationally representative data sets, like the CPS or SIPP. Although Moore, Stinson, and Welniak (1990) and Roemer (1999) conclude that income, especially labor earnings, are measured well in the CPS and SIPP, we believe that

several important questions remain with respect to income and employment measurement for low-income populations with national surveys.

- First, none of these studies, to our knowledge, focus on the reporting of income by disadvantaged, welfare-eligible, and/or welfare-prone populations.
- Second, as noted in Primus et al. (1999), participation in welfare programs is underreported in the CPS (and SIPP). Moreover, this underreporting appears to have increased over time. This is a troubling problem, especially as one looks to the future when TANF programs become state-specific, with different names.

Recommendation 1: We would like to see further work on the sources of antipoverty program underreporting and its origins in nationally representative survey data.

Plans are under way for some of the needed work. Joseph Hotz is a principal investigator on a project recently approved by the U.S. Census Bureau to match data from UI wage records and administrative data on AFDC/TANF participation for the California subsamples of several waves of the SIPP.³² This project should yield more recent information on both the welfare participation underreporting and income reporting issues. This study—or comparable ones matching SIPP with administrative data for the subsamples from other states—also may provide some insight into the impact of changes in family structure on income reporting for welfare leavers by exploiting the (limited) panel structure of the SIPP.

Further research is also needed on the use of UI wage records to measure the income of low-income and welfare-prone populations. Kornfeld and Bloom (1999) suggested that UI wage data and survey data produced similar estimates of the impact of a social program (i.e., JTPA-funded training programs) on earnings and employment, but their study also found that average earnings of JTPA-eligible individuals were consistently lower than those based on survey data. Furthermore, the study by Hill et al. (1999) also found that UI wage data produced substantially lower estimates of earnings than did tax return data for a welfare-based population drawn from the California AFDC caseload. Learning more

³²The other investigators on this project are David Card, Andrew Hildreth, and Michael Clune at UC-Berkeley and Robert Schoeni at RAND.

about the quality of this data source for measuring income is extremely important since UI wage data presumably will continue to be a core resource in state and local evaluations of the effects of welfare reform.

Several issues related to UI wage data seem to need further scrutiny. First, the study by Burgess, Blakemore, and Low (1998) raises important concerns about the “coverage” of UI and tax returns, particularly for the low-income population.

Recommendation 2: It would be extremely useful to follow the helpful lead of the various studies by Burgess and his colleagues to closely examine the coverage and trends in coverage of low-income populations with UI data. Such an examination could be aided by using a match of UI data with respondents in a national survey, such as the SIPP, so that one could learn more about the demographic characteristics of individuals (and households) that report labor market earnings on a survey that are not recorded in UI wage records data.

- States may be able to augment UI data used for evaluation of welfare reform by collecting supplemental information on the degree to which employers are designating workers as independent contractors. Additional work at the state level to assess the overall coverage of UI data would also be valuable.

Second, more work is needed to understand the extent to which UI wage data provide a misleading measure of the earnings available to low-income *households*. This problem arises in short- and long-term follow-up analyses of earnings for welfare samples drawn from state caseloads. One can use UI data to measure subsequent earnings for individuals in assistance units as long as they remain on welfare. However, as noted by Rolston (1999), one may not be able to accurately measure household income *after* assistance units leave the rolls because it is difficult to keep track of the identities of household members. The evidence provided in the Meyer and Cancian (1998) and Hill et al. (1999) studies suggests that this *may* be a serious problem.

Recommendation 3: To learn more about family well-being, it will be necessary to continue to rely on targeted, follow-up surveys to monitor samples of welfare leavers. Unfortunately, surveys are

expensive. We recommend that a pilot study be undertaken to devise a survey designed simply to obtain SSNs of other adults in a household, which can then be used to obtain UI wage earnings for these family members.

- State TANF agencies might analyze the methods used by their JTPA agencies to gather follow-up earnings data on terminees from their programs. Such follow-up assessments are required under JTPA, and many states have contracted with firms and/or universities to gather these follow-up data.
- Tax return data may also be useful to learn more about whether the discrepancies between UI wage data and income measures from tax returns noted in that study are the result of differences in family composition and the “composition” of income reported on tax returns.

A third issue relates to the possibility that wage earnings are missed because individuals move out of the state from which UI wage data are drawn or that workers earn part of their income in other states.

Again, comparisons of UI wage data with data from federal tax returns may help us to assess the importance of this problem and, more important, the biases that it imparts on measures of individual and household income. To learn more, it may be useful to take a closer look at what is known about the interstate mobility of disadvantaged and welfare-prone populations, such as the work done on movements of welfare populations in response to “welfare magnets,” as in Meyer (1999) and the citations therein, and the implications this mobility has for the coverage of low-income workers in UI data.

References

- Abraham, Katherine G., James R. Spletzer, and Jay C. Stuart. 1998. "Divergent Trends in Alternative Wage Series." In *Labor Statistics Measurement Issues*, edited by Haltiwanger, Manser and Topel. National Bureau of Economic Research Studies in Income and Wealth, vol. 60. Chicago: University of Chicago Press. Pp. 293–324.
- Baj, John, Sean Fahey, and Charles E. Trott. 1992. "Using Unemployment Insurance Wage-Record Data for JTPA Performance Management." Research Report 91-07, National Commission for Employment Policy. Chapter 4.
- Baj, John, Charles E. Trott, and David Stevens. 1991. "A Feasibility Study of the Use of Unemployment Insurance Wage-Record Data as an Evaluation Tool for JTPA: Report on Project Phase 1 Activities." National Commission on Employment Policy.
- Bane, Mary Jo, and David T. Ellwood. 1983. "The Dynamics of Dependence: The Routes to Self-Sufficiency." Prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. Cambridge, MA: Urban Systems Research and Engineering.
- Blakemore, Arthur E., Paul L. Burgess, Stuart A. Low, and Robert D. St. Louis. 1996. "Employer Tax Evasion in the Unemployment Insurance Program." *Journal of Labor Economics* 14: 210–230.
- Bogen, Karen, Meredith Lee, Julia Klein Griffiths, and Anne Polivka. 1997. "The Impact of Welfare Reform on the March CPS Income Supplement—Summary and Recommendations from Cognitive Interviews." Mimeo. U.S. Bureau of the Census and Bureau of Labor Statistics.
- Bogen, Karen. 1998. "Once Upon a Time, There Was Welfare Reform: Evaluating the New March CPS Welfare-Related Questions: Results from the 1998 Respondent Debriefing," Mimeo. U.S. Bureau of the Census.
- Bollinger, Christopher R., and Martin H. David. Forthcoming. "Estimation with Response Error and Non-Response: Food Stamp Participation in the SIPP." *Journal of Business and Economic Statistics*.
- Bound, John, and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* 9: 1–24.
- Burgess, Paul L., Arthur E. Blakemore, and Stuart A. Low. 1998. "Using Statistical Profiles to Improve Unemployment Insurance Tax Compliance." *Research in Employment Policy* 1: 243–271.
- Cancian, Maria, Robert Haveman, Thomas Kaplan, and Barbara Wolfe. 1999. "Post-Exit Earnings and Benefit Receipt among Those Who Left AFDC in Wisconsin." Special Report no. 75, Institute for Research on Poverty, University of Wisconsin–Madison.
- Cilke, Jim. 1998. "A Profile of Non-Filers." U.S. Department of Treasury, OTA Paper #78.
- Coder, John. 1992. "Using Administrative Record Information to Evaluate the Quality of the Income Data Collected in the SIPP." *Proceedings of Statistics Canada Symposium 92: Design and Analysis of Longitudinal Surveys*. Ottawa: Statistics Canada. Pp. 295–306.

- Coder, J., and Scoon-Rogers, L. S. 1996. "Evaluating the Quality of Income Data Collection in the Annual Supplement to the March Current Population Survey and the Survey of Income and Program Participation." SIPP Working Paper 96-04.
- Greenberg, David, and Harlan Halsey. 1983. "Systematic Misreporting and the Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment." *Journal of Labor Economics* 1: 380–407.
- Halsey, Harlan. 1978. "Validating Income Data: Lessons from the Seattle and Denver Income Maintenance Experiment." Proceedings of the Survey of Income and Program Participation Workshop, U.S. Department Health, Education and Welfare. Pp. (IV)21–(IV)51.
- Hill, Carolyn, V. J. Hotz, Charles Mullin, and John Karl Scholz. 1999. "EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload." Report submitted to the California Department of Social Services.
- Houseman, Susan N. 1999. "Flexible Staffing Arrangements: A Report on Temporary Help, On-Call, Direct-Hire Temporary, Leased, Contract Company, and Independent Contractors Employment in the United States."
<<http://www.dol.gov/dol/asp/public/futurework/conference/staffing/intro.htm>>
- Kornfeld, Robert, and Howard S. Bloom. (1999). "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17: 168–197.
- Marquis, K. H., and J. C. Moore. 1990. "Measurement Errors in SIPP Program Reports." Proceedings of the Bureau of the Census 1990 Annual Research Conference. Pp. 721–745.
- Meyer, Daniel R., and Maria Cancian. 1998. "Economic Well-Being Following an Exit from Aid to Families with Dependent Children." *Journal of Marriage and the Family* 60: 479–492.
- Meyer, Bruce D. 1999. "Do the Poor Move to Receive Higher Welfare Benefits." Mimeo. Northwestern University Economics Department.
- Moore, J., K. Marquis, and K. Bogen. 1996. "The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation." U.S. Bureau of the Census.
- Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak Jr. 1997. "Income Measurement Error in Surveys: A Review." Statistical Research Report, U.S. Bureau of the Census.
- Nelson, Charles T., Marc I. Roemer, Daniel H. Weinberg, and Edward J. Welniak Jr. 1998 "Fifty Years of United States Income Data from the Current Population Survey." Mimeo. Housing and Household Economics Statistics Division, U.S. Bureau of the Census.
- Polivka, Anne P. 1997. "Using Earnings Data from the Current Population Survey after the Redesign." Mimeo. Bureau of Labor Statistics.
- Polivka, Anne P. 1998. "Note on the Possible Effects of Welfare Reform on Labor Market Activities: What Can Be Gleaned from the March CPS." Mimeo. Bureau of Labor Statistics.

- Primus, Wendell, Lynette Rawlings, Kathy Larin, and Kathryn Porter. 1999. "The Initial Impacts of Welfare Reform on the Incomes of Single-Mother Families." Center for Budget and Policy Priorities.
- Rockefeller Institute of Government. 1999. "After Welfare: A Study of Work and Benefit Use after Case Closing in New York State." Revised interim report submitted to the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.
- Rodgers, Willard L., Charles Brown, and Greg J. Duncan. 1993. "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages." *Journal of the American Statistical Association* 88: 1208–1218.
- Roemer, Marc. 1999. "Reconciling March CPS Money Income with the National Income and Product Accounts: An Evaluation of CPS Quality." Mimeo. Income Statistics Branch, Bureau of the Census.
- Roemer, Marc. 2000. "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990–1996." Mimeo. Income Statistics Branch, Bureau of the Census.
- Rolston, Howard. 1999. "The Income of Former Welfare Recipients." Mimeo. Administration on Children and Families, U.S. Department of Health and Human Services.
- Sailer, Peter, and Michael Weber. 1998–1999. "The IRS Population Count: An Update." In *Turning Administrative Systems into Information Systems*, pp. 85–89. Available on the Web at <<http://ftp.fedworld.gov/pub/irs-soi/indpopct.pdf>>
- Scholz, John Karl. 1994. "The Earned Income Tax Credit: Participation, Compliance and Anti-Poverty Effectiveness." *National Tax Journal*, March, pp. 59–81.
- Shroder, Mark, and Marge Martin. 1996. "New Results from Administrative Data: Housing the Poor, or, What They Don't Know Might Hurt Somebody." Mimeo. Office of Policy Development and Research, U.S. Department of Housing and Urban Development.
- Smith, Jeffrey. 1997. "Measuring Earnings Levels among the Poor: Evidence from Two Samples of JTPA Eligibles." Mimeo. Department of Economics, University of Western Ontario.
- Stevens, David W., Liping Chen, and Jinping Shi, 1994. "The Use of UI Wage Records for JTPA Performance Management in Maryland." Mimeo. The Jacob France Center at the University of Baltimore.
- U.S. Department of Commerce, Bureau of the Census. 1998. "SIPP Quality Profile, 1998." SIPP Working Paper 230, 3rd Edition.
- Wisconsin Department of Workforce Development. 1999. "Differences between AFDC and W-W Leavers Survey Data for January-March 1998 and Wisconsin's UI Wage Records for 1998." DWD MEP Folio Brief 01-99.

Yen, W., and H. Nelson. 1996. "Testing the Validity of Public Assistance Surveys with Administrative Records: A Validation Study of Welfare Survey Data." Unpublished manuscript. May.